AIRBNB Case Study Methodology

Presentation 1 –

In the case study we have used Jupiter notebook to perform initial analysis of the data and Tableau for data analysis and visualization.

Step 1-
Initial Analysis using Jupiter Notebook:

- Data Set Used: AB_NYC_2019.csv
- Number of Rows: 48895
- Number of Columns: 16
- We removed the columns like Id, Name, Last Review which was not giving much information.

Step 2 - Data Wrangling:

- Checked the Duplicate rows in our dataset and no duplicate data were found.
- Checked the Null Values in our dataset. Columns like name, host-name, last review and review-per-month have null values.
- We've dropped the column name as missing values are less and dropping it won't have significant impact on analysis.
- Checked the formatting in our dataset.
- Identified and review outliers.

Data Analysis and Visualizations using Tableau:

We have used tableau to visualize the data for the assignment.

1) Top 10 Host: We identified the top 10 Host Ids, Host Name with count of Host Ids using the tree map.
2) Preferred Room type with respect to Neighborhood group: We created a pie chart for understanding the percentage of room type preferred w r t neighborhood group
3) For Variance of price with Neighborhood Groups:
We used a box and whisker's plot with Neighborhood Groups in Columns and Price in Rows.
We changed the Price from a Sum Measure to the median measure.
4) Average price of Neighborhood groups:
We created a bubble chart with Neighborhood Groups in Columns and Price columns in Rows.
We added the Neighborhood Groups to the colors Marks card to highlight the different neighborhood Groups in different colors. Also Put Avg price in Label.
5) Customer booking w r t minimum nights:
We created the bin for Minimum nights

The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighborhood group.
6) Popular Neighborhoods:
We took neighborhood in rows and sum of reviews in column and took neighborhood groups in color.
We used filter to show Top 20 neighbors as per the sum of reviews.
7) Neighborhood vs Availability:
We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighborhood group sorted by price.

Presentation 2 -

1) Room type with respect to Neighborhood group:
We created a pie chart for understanding the percentage of room type preferred w r t neighborhood group
We added Room Type to the colors Marks card to highlight the different Room Type in different colors and count of Host Id to the size
2) Customer Booking with respect to minimum nights:
We created the bin for Minimum nights
The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighborhood group.
3) Neighborhood vs Availability:
We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighborhood group sorted by price.
4) Price range preferred by Customers:
We have taken pricing preference based on volume of bookings done in a price range and no of Ids to create a bar chart. We have created bin for Price column with interval of $20.
5) Understanding Price variation w.r.t Room Type & Neighborhood:
We created Highlights Table chat by taking Room Type in rows & Neighborhood Group in column.
We took the average price in color Marks card to highlight the different Room Type
6) Price variation w r t Geography:
We used Geo location chart to plot neighborhood, neighborhood Group in map to show case the variation of prices across.
7) Popular Neighborhoods:
We took neighborhood in rows and sum of reviews in column and took neighborhood groups in color.
We used filter to show Top 20 neighbors as per the sum of reviews.
8) Tools used:
Data cleaning and preparation: Jupyter notebook – Python
Visualization and analysis: Tableau